



Kostyk, A., Zhou, W., Hyman, M. R. and Paas, L. (2021) Securing higher-quality data from self-administered questionnaires. *International Journal of Market Research*, (doi: [10.1177/14707853211057172](https://doi.org/10.1177/14707853211057172))

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/258880/>

Deposited on 12 November 2021

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Securing Higher-quality Data from Self-Administered Questionnaires

Alena Kostyk

Lecturer in Marketing, Adam Smith Business School
University of Glasgow, University Avenue, Glasgow G12 8QQ

Email: Alena.Kostyk@glasgow.ac.uk

ORCID: <https://orcid.org/0000-0003-2691-0603>

Wenkai Zhou

Assistant Professor of Marketing, College of Business
University of Central Oklahoma, Edmond OK 73034

Voice Phone: 405-974-2153

Email: wzhou2@uco.edu

ORCID: <https://orcid.org/0000-0002-5966-3621>

Michael R. Hyman

Distinguished Achievement Professor of Marketing, College of Business
New Mexico State University, Las Cruces, NM 88003-8001

Voice Phone: 575-522-8463

Email: michaelrhyman88011@gmail.com

ORCID: <https://orcid.org/0000-0001-6675-8808>

ResearcherID: E-3442-2018

Scopus Author ID: 7005893805

Leo Paas

Professor of Marketing, Business School
The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

Voice Phone: +64 9 373 7599

Email: leo.paas@auckland.ac.nz

ORCID <http://orcid.org/0000-0002-6611-3038>

© 2021 by Alena Kostyk, Wenkai Zhou, Michael R. Hyman, and Leo Paas

Securing Higher-quality Data from Self-Administered Questionnaires

Abstract

After highlighting the primary purposes and quality considerations of survey research, we briefly discuss previously recommended approaches for ensuring and improving data quality. Then, we introduce the articles in this special issue that feature novel solutions for improvement.

Will behavioural and ‘big’ data replace or complement survey data (Miller, 2017; Sturgis and Luff, 2020)? Seemingly, the answer is the latter. Recent studies suggest no decline in consumer survey usage (Sturgis and Luff, 2020). Unlike behavioural and ‘big’ data, the enduring benefits of survey methodology include investigating ‘why’ (i.e., tapping into underlying attitudes, beliefs, and values), offering patrons a channel for voicing their pleasures and concerns, assessing inherently subjective and thus indirectly observable marketing constructs (e.g., processing fluency, which is better captured by self-reports than proxy measures), and reducing privacy concerns via informed consent (Kostyk et al., 2021; SurveyMonkey, 2021). In addition, survey data can provide a more comprehensive perspective on consumer behaviour than transactional data and online data. In essence, surveys will continue to be a vital data source for academic and industry research (Miller, 2017).

Nevertheless, academic and applied marketing researchers will need to adapt their surveys to an evermore challenging respondent environment characterised by low participation rates, shortened attention spans, fraudulent responses, server farms, and ‘bot’ respondents (Bohannon, 2016; Couper, 2013; Dennis et al., 2020; Perkel, 2020). Insensitivity to this new data collection reality will produce inaccurate results that increasingly discredit assessments of public sentiment and behavioural intentions. For example, consumer surveys and political polls differ in purpose and scientific rigour, yet recent inaccurate political polling has caused eligible voters to question the accuracy of survey research (National Research Center, 2019). Regardless, survey researchers must strive to minimise reputational damage and optimize survey-data-based decisions by enhancing data quality.

Many marketing scholars have scrutinised survey data quality (Goodman et al., 2013; Peterson, 2001; Sears, 1986; Smith et al., 2016). We organised Table 1 around five types of currently recognised threats to survey data quality: (1) respondent recruitment, (2)

respondents' attentiveness and response accuracy, (3) respondents' engagement and interest, (4) data collection on new, tech-enhanced platforms, and (5) post-collection data handling. A survey's primary purpose determines these issues' severity. For example, most of these issues matter little for surveys meant to give voice, yet attentiveness and accuracy matter considerably for surveys focused on attitudes, motives, and beliefs.

----- Place Table 1 here -----

Survey researchers have proposed different solutions for each data quality threat based on the relevant emerging scholarship. Survey design-related solutions often focus on participant recruitment (Paas, Dolnicar, and Karlsson, 2018; Zhang and Conrad, 2014), as well as multi-method and modular survey designs (Bansal et al., 2017). Questionnaire design solutions revolve around the layout and the features of the data collection instrument itself. They include presentation of instructions (Brosnan, Babakhani, and Dolnicar, 2019), attention checks and warnings (Meade and Craig, 2012; Oppenheimer, Meyvis and Davinden, 2009; Paas and Morren, 2018), and novel question formats and gamification (Dolnicar, Grün, and Yanamandram, 2013).

Researchers must carefully select and implement these solutions depending on the survey's primary purpose. Although attention checks, gamification, surveytainment, and warnings can improve data quality for surveys designed to enhance marketing scholarship, in some circumstances these techniques may be inadvisable because they bias respondents' cognitive and affective processes (Abbey and Meloy, 2017; Kostyk, Zhou, and Hyman, 2019). Alternatively, these methods may be acceptable for benchmarking (e.g., longitudinal customer satisfaction surveys), forecasting, or population profiling surveys meant to accurately estimate demographic, socio-economic, and other population parameters. For such surveys, a representative sample is critical to securing accurate population estimates, and thus researchers can offer creative incentives to motivate potential participants.

Post-collection data handling can affect analytical results markedly because even a small percentage of untrustworthy data can significantly bias reported findings (Barnette, 1999; Bollen and Arminger, 1991; Lind and Zumbo, 1993, Liu and Zumbo, 2007). Again, the recommended methods differ depending on the survey's purpose. For example, data imputation techniques (e.g., mean substitution, listwise deletion) may boost population profiling or forecasting survey accuracy, and composite data quality indices can boost the accuracy of forecasting models. However, for surveys meant to advance marketing scholarship about consumer attitudes and beliefs, identifying and deleting careless or mischievous responses is more critical than data imputation (Hyman and Sierra, 2012; Perkel, 2020).

Misleading data can yield published reports that retard scientific progress or encourage harmful business actions. Some conventional wisdom about survey data collection, cleaning, and transformation may need revisiting, as many iconic survey research texts were largely or wholly written during the pre-internet and pre-social media era (Dillman, Smyth, and Christian, 2009; Payne, 1951). Marketing researchers require new approaches to ensure high-quality survey data collected across data collection modes. Hence, this special issue of the *International Journal of Market Research* is dedicated to novel approaches for improving data quality from self-administered questionnaires. It represents an opportunity for marketing researchers to share their insights about evaluating and enhancing survey data quality. The four articles herein address multiple challenges encountered by a survey researcher and offer recommendations for overcoming them.

Special Issue Overview

In "Data quality assurance: What marketing researchers absolutely need to remember," Moore, Harrison, and Hair propose a data quality assurance framework that emphasises survey-based research methodology and related data collection sources. The

framework presents step-by-step guidance about ‘best practices’ for ensuring data quality and improving research outcomes. Supported by the relevant literature, the proposed framework encourages a holistic approach to tackling data-related research challenges from initial conceptualisation to the communication of results. Although they strongly encourage preemptive measures to reduce data quality problems encountered during early research stages, the authors also note that rigorous data processing techniques would further ensure data quality and research integrity. In essence, researchers can confidently present more robust research results by proactively integrating data quality assurance mechanisms into the research design.

In “Enhancing self-administered questionnaire response quality using reminders,” Saunders and Kulchitsky present a novel approach (i.e., embedded code-of-conduct reminders) for prompting survey participants to respond more consistently with researchers’ integrity expectations. Improved consistency can encourage more systematic and ethical responses, reducing the likelihood of participants acting unethically and providing lower-quality data. The authors relate the Hunt-Vitell General Theory of Marketing Ethics to survey respondents’ teleological and deontological tendencies and code-of-conduct reminders’ effect on questionnaire responses.

Phillips argues in “Using examples to increase recall in self-administered questionnaires” that the conventional wisdom about including long and exhaustive lists of examples in self-administered questionnaires to help respondents recall infrequent and irregular events may backfire. Enlightened by part-set cuing theory, Phillips’ empirical study suggests truncated lists of examples that cue non-prototypical subcategories can aid recall while increasing questionnaire design efficiency. She also recommends aiding recall via examples that cue low-accessibility or non-prototypical subcategories (versus high-accessibility or prototypical subcategories). The underlying mechanisms that yield such

changes may be due to respondents' subcategory knowledge. Although the study does not directly tie these findings to improved data quality and accurate responses, it lays the foundation for exploring possible best practices.

Roster's "Using memes in online surveys to engage and motivate respondents" relies on a comprehensive literature review and a mixed-methods approach to examine the effect of Internet memes on survey participants' engagement levels and response quality. Although the quantitative experimental findings were indecisive, deconstruction of participants' open-ended qualitative responses suggests that embedded memes make questionnaires more entertaining and less burdensome to complete.

Collectively, the four articles acknowledge the vital role that self-administered questionnaires play in marketing and consumer behaviour research. Consistent with this special issue's mission, the authors suggest novel and likely effective approaches for improved data collecting and processing. Notwithstanding often overblown claims about insights gleanable from 'big data', survey research will remain vital to studying consumers' needs, perceptions, attitudes, and behavioural intentions. Hence, these articles suggest underexplored survey research approaches that should improve research reliability.

Table 1

Threats to Survey Data Quality and Possible Solutions

Research Purpose	Threats and Possible Solutions				
	Respondents' Recruitment	Respondents' Attentiveness and Response Accuracy	Respondents' Engagement and Interest	Data Collection on New, Tech-enhanced Platforms	Post-collection Data Handling
Uncover 'why': Attitudes, motives, beliefs	<u>Issue severity: Moderate</u> Survey design solutions: <ul style="list-style-type: none"> Filter undesirable respondents before data collection based on selected personal characteristics (e.g., tendency for speeding, Paas, Dolnicar, and Karlsson, 2018; Zhang and Conrad, 2014) 	<u>Issue severity: High</u> Survey design solutions: <ul style="list-style-type: none"> Supplement survey data via multi-method research designs (e.g., focus groups and depth interviews) Questionnaire design solutions: <ul style="list-style-type: none"> Better presentation of questionnaire instructions (Brosnan, Babakhani, and Dolnicar, 2019) 	<u>Issue severity: High</u> Questionnaire design solutions: <ul style="list-style-type: none"> Use novel question formats and gamification (Dolnicar, Grün, and Yanamandram, 2013; cf. Guin, Baker, Mechling, and Ruyle, 2012) 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> Leverage web-based platforms (Casler, Bickel, and Hackett, 2013; Dodou and de Winter, 2014; Peterson et al., 2017) Use high-tech features such as embedded media and augmented or virtual reality (e.g., video/animation versus text for posing questions) 	<u>Issue severity: High</u> Solutions: <ul style="list-style-type: none"> Identify and delete careless or mischievous responses (Hyman and Sierra, 2012; Meade and Craig, 2012)
Population profiling and forecasting	<u>Issue severity: High</u> Survey design solutions: <ul style="list-style-type: none"> Offer novel incentives for survey participation based on cooperation norms (Kropf and Blair, 2005) Use novel forms of participant motivation (e.g., charity/political cause donation; Groves, Singer, and Corning, 2000; 	<u>Issue severity: High</u> Survey design solutions: <ul style="list-style-type: none"> Use modular survey design (Bansal et al., 2017) Questionnaire design solutions: <ul style="list-style-type: none"> Improve attention checks and warnings (Paas and Morren, 2018) 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> Use novel question formats and gamification Use interactive and non-interactive entertainment breaks (Kostyk, Zhou, and Hyman, 2019) 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> Leverage web-based platforms 	<u>Issue severity: High</u> Solutions: <ul style="list-style-type: none"> Improve data imputation techniques Identify and delete careless or mischievous responses Develop composite data quality indices

Research Purpose	Threats and Possible Solutions				
	Respondents' Recruitment	Respondents' Attentiveness and Response Accuracy	Respondents' Engagement and Interest	Data Collection on New, Tech-enhanced Platforms	Post-collection Data Handling
	cf. Singer and Couper, 2008)				
Benchmarking	<u>Issue severity: High</u> Survey design solutions: <ul style="list-style-type: none"> • Use novel incentives for survey participation based on cooperation norms, charity/political cause donations, etc. 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> • Improve attention checks and warnings 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> • Use interactive and non-interactive entertainment breaks 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> • Leverage web-based platforms 	<u>Issue severity: High</u> Solutions: <ul style="list-style-type: none"> • Identify and delete careless or mischievous responses
Giving voice to participants	<u>Issue severity: Low</u> Solutions: <ul style="list-style-type: none"> • None needed 	<u>Issue severity: Low</u> Solutions: <ul style="list-style-type: none"> • None needed 	<u>Issue severity: Low</u> Solutions: <ul style="list-style-type: none"> • None needed 	<u>Issue severity: Low</u> Solutions: <ul style="list-style-type: none"> • None needed 	<u>Issue severity: Moderate</u> Solutions: <ul style="list-style-type: none"> • Identify and delete mischievous responses
Enhancing knowledge	<u>Issue severity: Moderate</u> Survey design solutions: <ul style="list-style-type: none"> • Use novel incentives for survey participation based on cooperation norms, charity/political cause donations, etc. 	<u>Issue severity: High</u> Survey design solutions: <ul style="list-style-type: none"> • Supplement survey data via multi-method research designs • Use modular survey design Questionnaire design solutions: <ul style="list-style-type: none"> • Improve attention checks and warnings 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> • Use novel question formats and gamification • Use interactive and non-interactive entertainment breaks 	<u>Issue severity: Moderate</u> Questionnaire design solutions: <ul style="list-style-type: none"> • Leverage web-based platforms • Use high-tech features such as embedded media and augmented or virtual reality (e.g., use of video/animation versus text for posing questions) • Use tech-enhanced methods for assessing respondents' attentiveness and engagement (e.g., eye-tracking; Brosnan, 	<u>Issue severity: High</u> Solutions: <ul style="list-style-type: none"> • Identify and delete careless or mischievous responses • Identify novel data quality indicators for post-hoc data cleaning (Aust et al., 2013) • Develop composite data quality indices

Research Purpose	Threats and Possible Solutions				
	Respondents' Recruitment	Respondents' Attentiveness and Response Accuracy	Respondents' Engagement and Interest	Data Collection on New, Tech-enhanced Platforms	Post-collection Data Handling
				Babakhani, and Dolnicar, 2019)	

References

- Abbey JD and Meloy MG (2017) Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management* 53: 63-70. DOI: 10.1016/j.jom.2017.06.001
- Aust F, Diedenhofen B, Ullrich S, and Musch J (2013) Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods* 45(2): 527-535.
- AAPOR Report on non-probability sampling (2013) Available at: <https://www.aapor.org/Education-Resources/Reports/Non-Probability-Sampling.aspx> (accessed 27 July 2021).
- AAPOR report on evaluating survey quality in today's complex environment (undated) Available at: <https://www.aapor.org/Education-Resources/Reports/Evaluating-Survey-Quality.aspx> (accessed 27 July 2011).
- Bansal HS, Eldridge J, Halder A, Knowles R, Murray M, Sehmer L, and Turner D (2017) Shorter interviews, longer surveys: Optimising the survey participant experience while accommodating ever expanding client demands. *International Journal of Market Research* 59(2): 221-238. DOI: 10.1177/1470785319870622a
- Barnette JJ (1999) Nonattending respondent effects on interval consistency of self-administered surveys: A Monte Carlo simulation study. *Educational and Psychological Measurement* 59(1): 38-46. DOI: 10.1177/0013164499591003
- Bohannon J (2016) Many surveys, about one in five, may contain fraudulent data. *Science*. Available at: <https://www.sciencemag.org/news/2016/02/many-surveys-about-one-five-may-contain-fraudulent-data> (accessed 18 August 2021).
- Bollen, KA and Arminger G (1991) Observational residuals in factor analysis and structural equation models. *Sociological Methodology* 21: 235-262. DOI: 10.1073/pnas.1010661108

- Brosnan K, Babakhani N, and Dolnicar S (2019) “I know what you’re going to ask me”: Why respondents don’t read survey questions. *International Journal of Market Research* 61(4): 366-379. DOI: 10.1177/1470785318821025
- Casler K, Bickel L, and Hackett E (2013) Separate but equal? A comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior* 29(6): 2156-2160. DOI: 10.1016/j.chb.2013.05.009
- Couper MP (2013) Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods* 7(3): 145-156. DOI: 10.18148/srm/2013.v7i3.5751
- Dennis SA, Goodson BM and Pearson CA (2020) Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. *Behavioral Research in Accounting* 32(1): 119-134. DOI: 10.2308/bria-18-044
- Dillman DA, Smyth JD, and Christian LM (2009) *Internet, Mail, and Mixed-mode Surveys: The Tailored Design Method (3rd ed.)*. New York, NY: Wiley.
- Dodou D and de Winter JC (2014) Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior* 36: 487-495. DOI: 10.1016/j.chb.2014.04.005
- Dolnicar S, Grün B, and Yanamandram V (2013) Dynamic, interactive survey questions can increase survey data quality. *Journal of Travel & Tourism Marketing* 30(7): 690-699. DOI: 10.1080/10548408.2013.827546
- ESOMAR World Research (2011) ESOMAR guideline for online research. Available at: <https://ana.esomar.org/documents/esomar-guideline-for-online-research-> (accessed 27 July 2021).

- ESOMAR World Research (2012) 28 questioners to help buyers of online samples. Available at: <https://www.esomar.org/what-we-do/code-guidelines/28-questions-to-help-buyers-of-online-samples> (accessed 27 July 2021).
- ESOMAR World Research (2015) ESOMAR/GBRN guideline for online sample quality. Available at: <https://ana.esomar.org/documents/esomar-grbn-guideline-for-online-sample-quality-> (accessed 27 July 2021).
- Fleischer A, Mead AD, and Huang J (2015) Inattentive responding in MTurk and other online samples. *Industrial and Organisational Psychology* 8(2): 196-202. DOI: 10.1017/iop.2015.25
- Guin TDL, Baker R, Mechling J, and Ruyle E (2012) Myths and realities of respondent engagement in online surveys. *International Journal of Market Research* 54(5): 613-633. DOI: 10.2501/IJMR-54-5-613-633
- Goodman JK, Cryder CE, and Cheema A (2013) Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making* 26(3): 213-224. DOI:10.1002/bdm.1753
- Groves RM, Singer E, and Corning A (2000) Leverage-saliency theory of survey participation: Description and an illustration. *The Public Opinion Quarterly* 64(3): 299-308. DOI: 10.1086/317990
- Hyman, MR and Sierra JJ (2012) Adjusting self-reported attitudinal data for mischievous respondents. *International Journal of Market Research* 54(1): 129-145. DOI: 10.2501/IJMR-54-1-129-145
- Kostyk A, Zhou W, and Hyman MR (2019) Using surveytainment to counter declining survey data quality. *Journal of Business Research* 95: 211-219. DOI: 10.1016/j.jbusres.2018.10.024

- Kostyk A, Leonhardt JM, and Niculescu M (2021) Processing fluency scale development for consumer research. *International Journal of Market Research* 63(3): 353-367.
- Kropf ME and Blair J (2005) Eliciting survey cooperation: Incentives, self-interest, and norms of cooperation. *Evaluation Review* 29(6): 559-575. DOI: 10.1177/1470785319877137
- Lind JC and Zumbo BD (1993) The continuity principle in psychological research: An introduction to robust statistics. *Canadian Psychology* 34(4): 407-414. DOI: 10.1037/h0078861
- Liu Y and Zumbo BD (2007) The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Visual analogue scales. *Educational and Psychological Measurement* 67(4): 620-634. DOI: 10.1177/0013164406296976
- Meade AW and Craig SB (2012) Identifying careless responses in survey data. *Psychological Methods* 17(3): 437-455. DOI: 10.1037/a0028085
- Miller PV (2017) Is there a future for surveys? *Public Opinion Quarterly* 81(S1): 205-212. DOI: 10.1093/poq/nfx008
- National Research Center (2019) What does the future of survey research look like? Available at: <https://blog.polco.us/future-survey-research-look-like> (accessed 27 July 2021).
- Oppenheimer, DM, Meyvis T, and Davidenko, N. (2009) Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4): 867-972. DOI:10.1016/J.JESP.2009.03.009
- Paas LJ, Dolnicar S, and Karlsson L (2018) Instructional manipulation checks: A longitudinal analysis with implications for MTurk. *International Journal of Research in Marketing* 35(2): 258-269. DOI: 10.1016/j.ijresmar.2018.01.003

- Paas LJ and Morren M (2018) Please do not answer if you are reading this: Respondent attention in online panels. *Marketing Letters* 29(1): 13-21. DOI: 10.1007/s11002-018-9448-7
- Payne SL (1951) *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.
- Perkel JM (2020) Mischief-making bots attacked my scientific survey. *Nature* 579(7798): 461-2. DOI: 10.1038/d41586-020-00768-0
- Peterson RA (2001) On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research* 28(3): 450-461. DOI: 10.1086/323732
- Peterson G, Griffin J, LaFrance J, and Li J (2017) Smartphone participation in web surveys. In: *Total survey error in practice* (eds. PP Beimer et al.), pp. 203-233. New York, NY: Wiley.
- Sears DO (1986) College sophomores in the laboratory: Influences of a narrow database on social psychology's view of human nature. *Journal of Personality and Social Psychology* 51(3): 515-530. DOI: 10.1037/0022-3514.51.3.515
- Singer E and Couper MP (2008) Do incentives exert undue influence on survey participation? Experimental evidence. *Journal of Empirical Research on Human Research Ethics* 3(3): 49-56. DOI: 10.1525/jer.2008.3.3.49
- Smith SM, Roster CA, Golden LL, and Albaum GS (2016) A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research* 69(8): 3139-3148. DOI: 10.1016/j.jbusres.2015.12.002
- Sturgis P and Luff R (2020) The demise of the survey? A research note on trends in the use of survey data in the social sciences, 1939 to 2015. *International Journal of Social Research Methodology*. DOI: 10.1080/13645579.2020.1844896

- SurveyMonkey (2021) Why surveys? How survey research and methodology make a difference. Available at: <https://www.surveymonkey.co.uk/mp/why-survey-understanding-survey-methodology/> (accessed 27 July 2021).
- Van Herk H, Poortinga YH, and Verhallen TM (2004) Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology* 35(3): 346-360. DOI: 10.1177/0022022104264126
- Zhang C and Conrad F (2014) Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods* 8(2): 127-135. DOI: 10.18148/srm/2014.v8i2.5453